Cardiovascular Disease Prediction using Machine Learning

Members : Aashna Shroff & Anushka Tare

Roll nos : 53 & 60 (BE 2)

Project Guide : Prof. Nida Parkar

Atharva college of Engineering

Problem Definition

- Heart disease is one of the most significant causes of mortality in the world today.
- Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis.
- Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry.
- This project proposes a prediction model to predict whether a people have a heart disease or not.

Introduction

- Heart is one of the most extensive and vital organ of human body so the care of heart is essential.
- Most of diseases are related to heart so the prediction about heart diseases is necessary and for this purpose comparative study needed in this field, today most of patient are died because their diseases are recognized at last stage due to lack of accuracy of instrument so there is need to know about the more efficient algorithms for diseases prediction.
- As the definition of machine learning, it learns from the natural phenomenon, natural things so in this project we uses the biological parameter as testing data such as cholesterol, blood pressure, sex, age, etc. and on the basis of these, comparison is done in the terms of accuracy of algorithms.

Review of Literature

- Lot of work has been carried out to predict heart disease using UCI Machine Learning dataset. Different levels of accuracy have been attained using various data mining techniques which are explained as follows.
- Research was carried out to study Decision Tree, KNN and K-Means algorithms that can be used for classification and their accuracy were compared.
- This research concludes that accuracy obtained by Decision Tree was highest further it was inferred that it can be made efficient by combination of different techniques and parameter tuning.
- The base paper being considered makes an attempt to detect these heart diseases at early stage to avoid disastrous consequences using different algorithm methods like Decision Tree, Naive Bayes, Logistic Regression, Random Forest, Heart Disease Prediction.

Proposed Solution

- **Data Preprocessing:**

  The first step is to get the dataset and then clean the data and make it viable to use.

- **Divide the dataset:**

  Then we divide the dataset into training dataset and testing dataset.

- **Features of the dataset:**

  Our data has features such as age, gender, height, weight, blood pressure, cholesterol, glucose level, smoker or not, etc.

  We use these data features by putting them into the machine learning algorithms and predicting results if a person will have cardiovascular disease or not.

Proposed Solution

- We use four machine learning algorithms to predict the results and the performance of the algorithms are compared to know which algorithm works best.
- **Naive Bayes:**

  Naive Bayes is a probabilistic classifier based on Bayes Theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

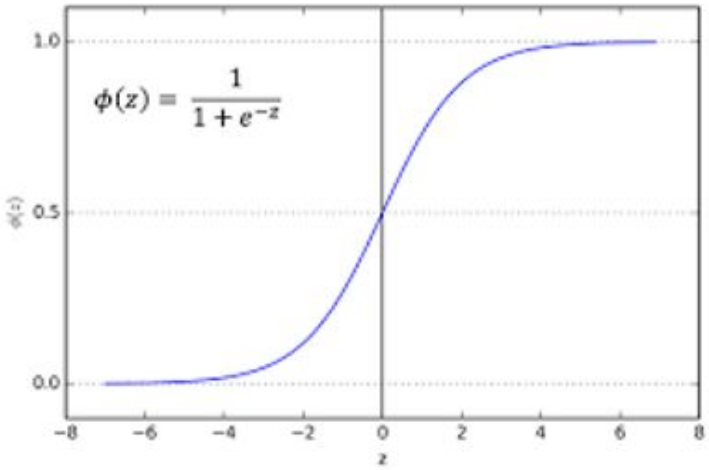
  Using Bayes theorem, we can find the probability of A happening, given that B has occurred. The assumption made here is that the predictors/features are independent.

- **Logistic Regression:**

  Logistic regression is basically a supervised classification algorithm.In a classification problem, the target variable(or output), y, can take only discrete values for given set of features(or inputs), X.

Proposed Solution

- Logistic regression models the data using the sigmoid function.



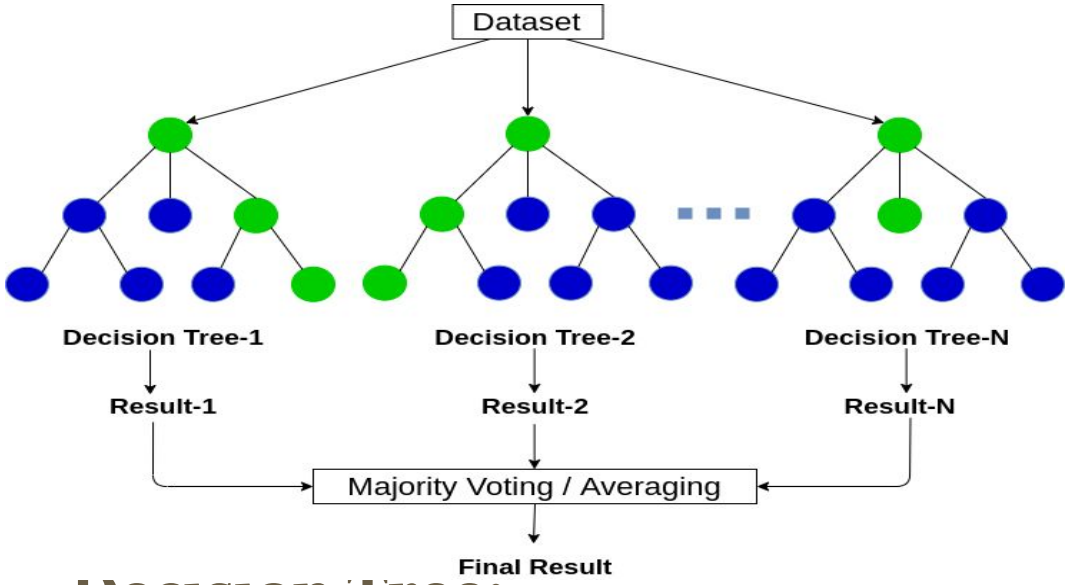
- **Random Forest:**

  Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

  The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result.

Proposed Solution

- **Random Forest:**

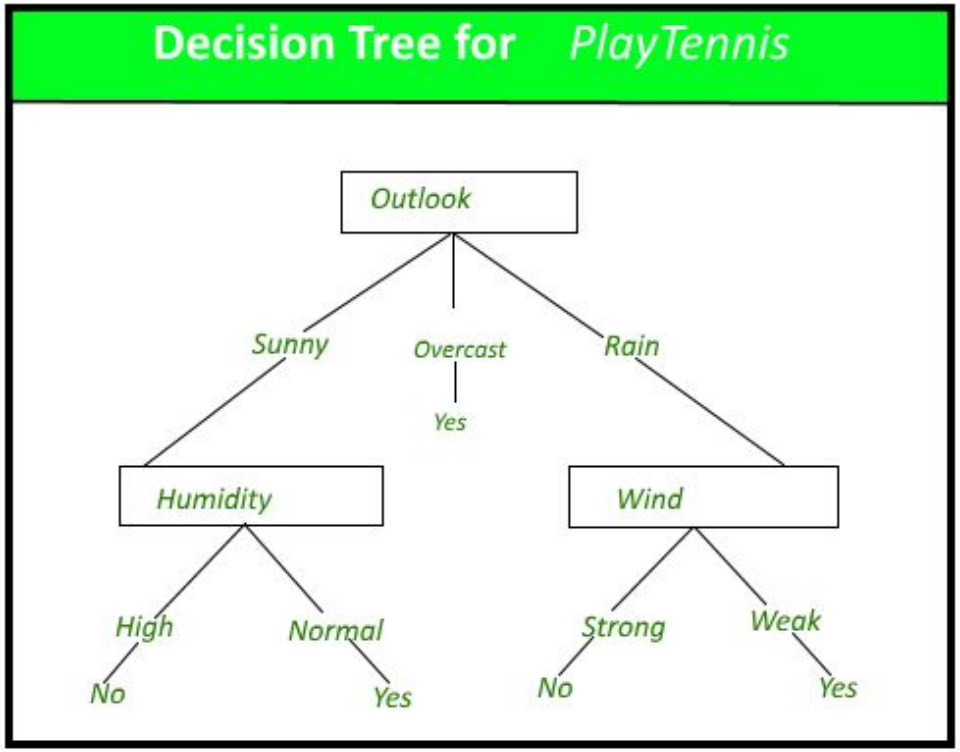


- **Decision Tree:**

  A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

Proposed Solution

- **Decision Tree:**



Proposed Solution

- **Predicting Accuracy:**

  Accuracy of the algorithms are depends on four values namely true positive(TP), false positive(FP), true negative(TN) and false negative(FN).

  Accuracy= (FN+TP) / (TP+FP+TN+FN).

  The numerical value of TP, FP, TN, FN defines as:

  TP= Number of person with heart diseases.

  TN= Number of person with heart diseases and no heart diseases.

  FP= Number of person with no heart diseases.

  FN= Number of person with no heart diseases and with heart diseases .

Feasibility

1.1.1 Operational Feasibility

Operational feasibility is the measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.

1.1.2 Technical Feasibility

Technical Feasibility, includes the development of the machine learning model. This model is developed fully in Python IDE/ Jupyter notebooks, thus there is no additional technical requirement. Also, as far as the This working model can be uploaded on Github or AWS cloud and can be deployed from here for further use.

1.1.3 Economic Feasibility

Economic feasibility is the cost and logistical outlook for this model. By using this model we can further avoid the expenses of required pathological services. This model can be deployed to use in facilities like hospitals, military, research centers for greater analysis of a patient's medical background.

Technological Dependencies

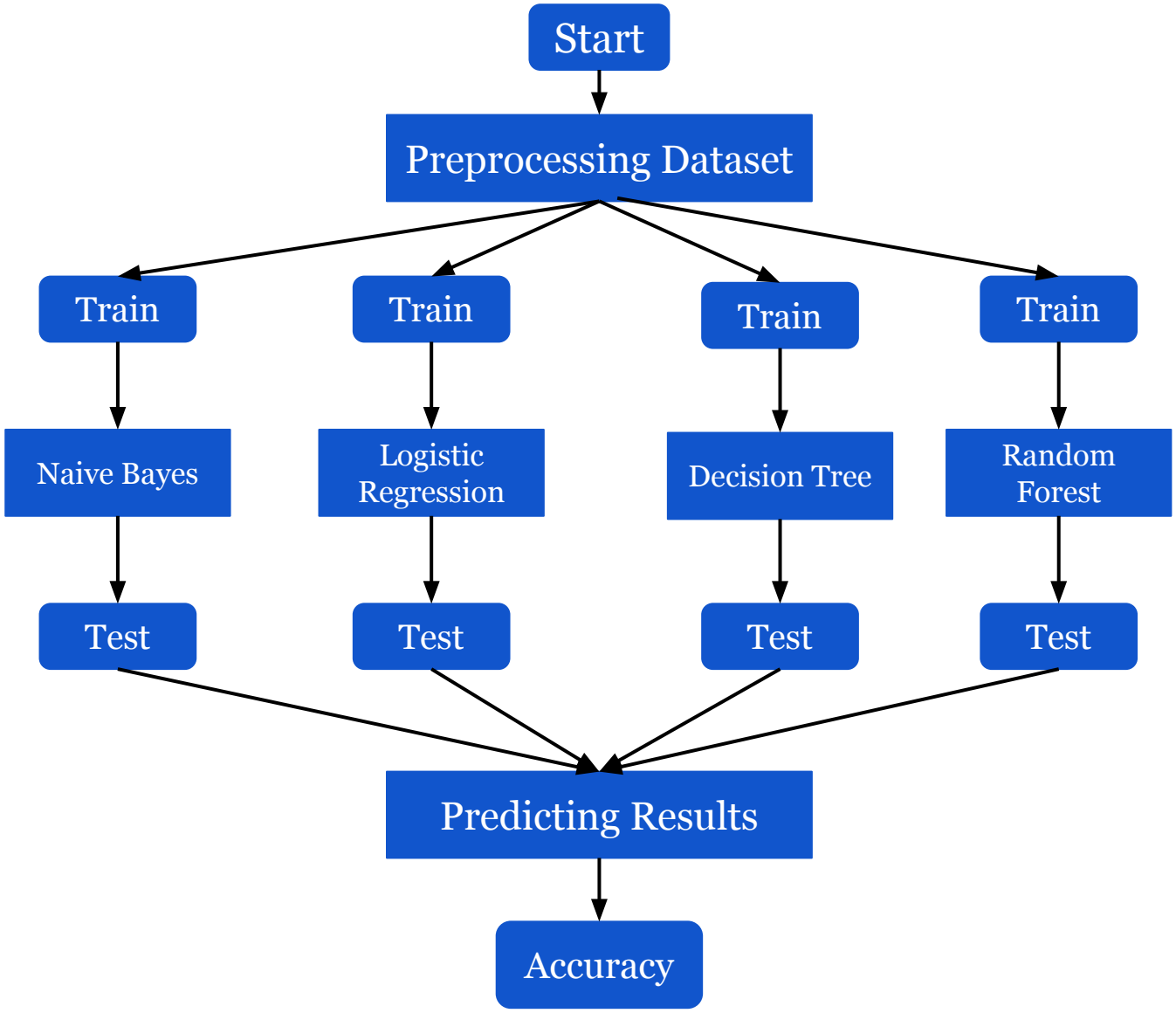
For System Development:

- Language : Python
- Platform: Jupyter Notebook, Python IDE
- Libraries: NumPy, SciPy, Matplotlib, Pandas, Sklearn etc
- MongoDB
- Apache Spark
- Flask
- ElasticSearch

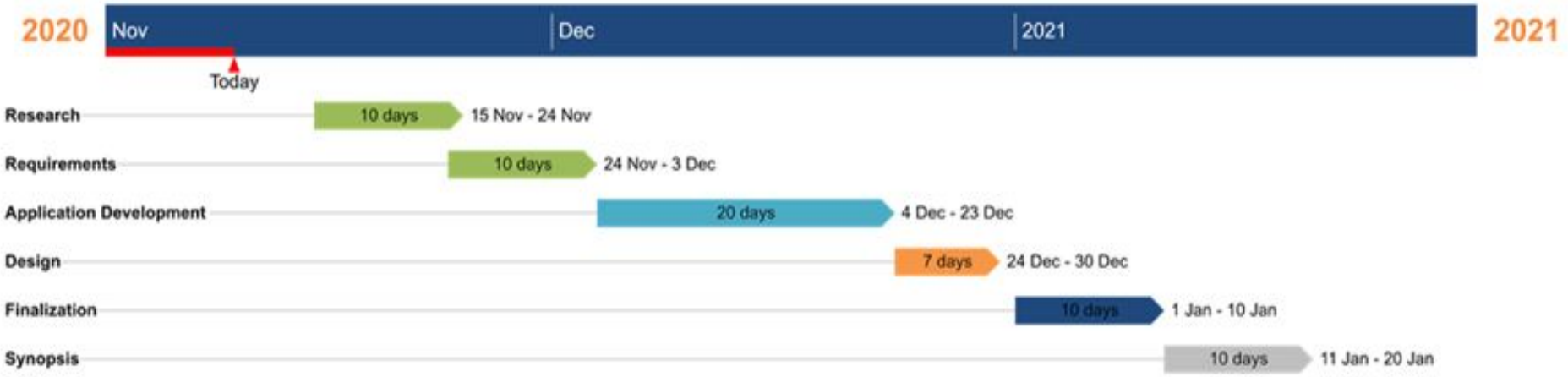
For System Deployment:

- AWS Cloud
- Github

Design Of Proposed System



Project Schedule



References

- Heart Disease Prediction using Machine Learning(Apurb Rajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli)https://www.ijert.org/heart-disease-prediction-using-machine-learning
- Heart Disease Prediction Using Machine Learning Algorithms(Publisher: IEEE Archana Singh; Rakesh Kumar)https://ieeexplore.ieee.org/abstract/document/9122958
- Prediction of Heart Disease Using Machine Learning Algorithms(Publisher: Science Publishing Corporation) https://www.researchgate.net/publication/325116774_Heart_disease_prediction_using_machine_learning_techniques_A_survey